DOCUMENT RESUME

ED 209 788 EC 140 098

AUTHOR Yarnall, Gary Dean; Carlton, Glenn R.
TITLE Guidelines and Manual of Tests for Educators Interested in the Assessment of Handicapped Children.
INSTITUTION International Research Inst., Austin, Tex.
PUB DATE 81
NOTE 114p.

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS *Disabilities; *Educational Diagnosis; Elementary Secondary Education; Resource Materials; *Student Evaluation; Test Reviews; *Tests

ABSTRACT

The manual contains reviews of 29 assessment tools and tests useful in testing physically, sensorially, or severely handicapped children. An introduction discusses the limitations of standardized intelligence tests, difficulties encountered in testing handicapped children, guidelines for testing, and specific testing suggestions for children who are blind/visually handicapped, deaf/hearing impaired, multihandicapped, orthopedically handicapped, or severely and profoundly handicapped. The tests are listed alphabetically. Reviews are in alphabetical order with each review containing information on purpose, organization, administration, technical data, limitations, and general considerations. The following tests are reviewed: Adaptive Behavior Scale, Bayley Scales of Infant Development, The Bender Gestalt Test, Boehm Test of Basic Concepts, Callier-Azusa Scale, The Columbia Mental Maturity Scale, Denver Developmental Screening Test, Gesell Developmental Schedules, Goodenough-Harris Draw-A-Man Test, Haptic Intelligence Scale, Hiskey-Nebraska Test of Learning Aptitude, The Leiter International Performance Scale, Maxfield-Buchholz Social Maturity Scale for Blind Preschool Children, Michigan Manual for the Assessment of Deaf-Blind, Multiple Handicapped Children, Ontario School Ability Examination, Peabody Individual Achievement Test, Peabody Picture Vocabulary Test, Pictorial Test of Intelligence, Porteus Maze Test, Preschool Attainment Record, Raven's Progressive Matrices, Santa Clara Inventory of Developmental Tasks, Sequential Tests of Educational Progress, Slosson Intelligence Test, Stanford Achievement Test, Stanford-Binet Intelligence Scale--1972 Norms Edition, Vineland Social Maturity Scale, and Wechsler Intelligence Scale for Children-Revised. (DB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED209788

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

GUIDELINES AND MANUAL OF TESTS FOR EDUCATORS INTERESTED IN THE ASSESSMENT OF HANDICAPPED CHILDREN

by

Gary Dean Yarnall, Ed.D.

and

Glenn R. Carlton, M.A.

Published by

International Research Institute
Post Office Box 3318
Austin, Texas 78764

Reviewed by

South Central Regional Center for
Services to Deaf-Blind Children
2930 Turtle Creek Plaza
Suite 207
Dallas, Texas 75219

1981

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Sharon S. Ebert
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

EC140078

Guidelines and Manual of Tests for Educators Interested in the Assessment of Handicapped Children

Gary Dean Yarnall
Louisiana State Department
of
Special Education
Baton Rouge, Louisiana

and

Glenn R. Carlton
Southwestern City Schools
Columbus, Ohio

January, 1981

NOTICE: Regard this as written permission to copy, use, and disseminate this at your discretion.

Gary Dean Yarnall

INDEX

	Page
Using This Manual	1
Introducation	2
Difficulties Encountered in Testing Handicapped Children	4
Guidelines for Testing	4
Blind/Visually Handicapped Children	7
Deaf Children/Hearing Impaired Children	10
Multihandicapped Children	11
Orthopedically Handicapped Children	13
Severely and Profoundly Handicapped Children	14
Review of 29 Assessment Tools/Tests	16
ADAPTIVE BEHAVIOR SCALE (ABS)	16
BAYLEY SCALES OF INFANT DEVELOPMENT	19
THE BENDER GESTALT TEST	23
BOEHM TEST OF BASIC CONCEPTS	26
CALLIER-AZUSA SCALE	28
THE COLUMBIA MENTAL MATURITY SCALE	31
DENVER DEVELOPMENTAL SCREENING TEST	35
GESELL DEVELOPMENTAL SCHEDULES	38
GOODENOUGH-HARRIS DRAW-A-MAN TEST	40
HAPTIC INTELLIGENCE SCALE	43
HISKEY-NEBRASKA TEST OF LEARNING APTITUDE	46

Page

THE LEITER INTERNATIONAL PERFORMANCE SCALE . . . 51

MAXFIELD-BUCHHOLZ SOCIAL MATURITY SCALE FOR BLIND PRESCHOOL CHILDREN 56

MICHIGAN MANUAL FOR THE ASSESSMENT OF DEAF-BLIND, MULTIPLE HANDICAPPED CHILDREN 59

ONTARIO SCHOOL ABILITY EXAMINATION 61

PEABODY INDIVIDUAL ACHIEVEMENT TEST 64

PEABODY PICTURE VOCABULARY TEST 68

PICTORIAL TEST OF INTELLIGENCE 71

PORTEUS MAZE TEST 74

PRESCHOOL ATTAINMENT RECORD (PAR) 77

RAVEN'S PROGRESSIVE MATRICES 80

SANTA CLARA INVENTORY OF DEVELOPMENTAL TASKS . . 83

SEQUENTIAL TESTS OF EDUCATIONAL PROGRESS (STEP) 86

SLOSSON INTELLIGENCE TEST 88

STANFORD ACHIEVEMENT TEST 90

STANFORD-BINET INTELLIGENCE SCALE: 1972 NORMS EDITIONS 93

VINELAND SOCIAL MATURITY SCALE 98

WECHSLER INTELLIGENCE SCALE FOR CHILDREN-- REVISED 101

References 106

Using This Manual

This manual describes many tests that are designed or may be adapted for the assessment of children with sensory, physical, or mental handicaps. The tests are arranged in alphabetical order and provide information concerning the various handicapping conditions for which they are applicable. In addition to using these formal devices, the examiner can gather information by interviewing the parents, and reviewing medical and educational records. Evaluation depends on rapport, observation, and time, plus accuracy of information supplied by those who accompany the child (Donlon, 1970).

Once the data have been collected, they should be arranged into a developmental sequence according to the areas assessed. Certain sequences of behavior are commonly seen at a particular stage of development. By indicating the age level where a child is performing a specific ability, the skills that have been mastered, and those the child is ready to learn, the examiner will provide useful information for placement and planni

Introduction

As a result of P.L. 94-142, many more children with handicaps of one sort or another are entering regular public schools. While the inclusion of severely handicapped children in public schools has been applauded as a major revolution in special education, as with any change of this magnitude, it is causing disruptions. One of the professional groups most concerned with the changes brought about by P.L. 94-142 is school psychologists. Most school psychologists have had little experience with severely impaired children, they are understandably concerned about the special skills and techniques they need to provide adequate services.

The psychologist plays a key role in carrying out the mandates of the new law. One of the requirements of the legislation is for a multifactored evaluation of all children before they are placed in special programs. For a number of years now, school psychologists, teachers, and other educators have recognized the pitfalls of placing children in special classes on the basis of a single test. Because intelligence tests measure ability indirectly, inferences made on the basis of a single IQ score are not necessarily valid. Standardized intelligence tests measure only a

child's performance on tasks that require mental work plus the amount of learning the child has acquired prior to the assessment. The validity of an intelligence test can be measured by correlating the test items and the total test score with achievement test performance.

These limitations of standardized intelligence tests and the recognition that other measures of individual differences are important to educational programming led to the requirement for a multifactored evaluation. Although there are no firm guidelines as to what must be included in factored assessment, most educators would include tests of fine and gross motor skills, selfcare skills, speech and language development, cognitive abilities, social-adaptive behavior, and social-emotional adjustment. Data gathered in these areas can be used for making educational decisions and will provide information that can form the basis for the legally required Individualized Educational Plan (IEP).

Faced with the challenge of providing a multifactored assessment for a new population, school psychologists may be wondering where to begin. This manual describes many of the formal, more frequently used, commercially available tests.

Difficulties Encountered in Testing Handicapped Children

The techniques and skills used to test children with sensory and physical handicaps are the same as those used with normal children, but it is more demanding to apply them. Normal children often require little or no encouragement and see the test situation as a challenge. Seriously handicapped children may be somewhat less involved in the testing process. They have less experience with novel situations, and frequently have habits that interfere with testing. There are several general problems that often come up. First, the time requirements of some tests may pose difficulties. Second, communication problems may make it difficult or impossible for a child to respond to items and for the tester to score the responses. Third, the handicapped child may become easily fatigued. Often they have little experience with prolonged period of work. Finally, there may be difficulty in establishing rapport and sustaining motivation.

Guidelines for Testing

Before actually testing a handicapped child, the examiner will need to determine the degree to which the child can respond to the test items. To gather this critical information, the examiner will need to spend a considerable period of time observing the child as he interacts with significant adults in familiar surroundings.

If the examiner spends some time watching the child, he or she may be able to avoid some situations that might otherwise hinder him or her from establishing rapport.

When working with children, it is also important to keep in mind your long-range and short-term objectives. This is especially true with assessment. Although many people consider only the final score, some especially critical information can be gathered throughout the evaluation process. This is true for all children--not just the handicapped.

One important goal of any evaluation process is to gather as much information about the child as possible so as to have a better understanding of him. It is important to recognize that the items a child cannot perform are as important as those items he can complete correctly. Additionally, how the child interacts with the materials reveals much about him--whether he is curious, easily frustrated, needs to be in control, whatever. These are all clues to how the child responds to life. These behaviors may not necessarily affect the child's overall score, but knowing about them provides invaluable information for educational programming. You will also want to make the session as successful and rewarding as possible.

Assessing a multiply or severely handicapped child presents a unique challenge. One of the simplest techniques is to involve at least two persons in the assessment. This

will enable one person to record responses and behaviors as well as sequence materials, while the other person works directly with the child. In this situation the session can flow better and the assessor is less likely to lose the child's attention. During the first meeting with the child, one team member can interview the parent, while the other is getting to know the child and the child is becoming accustomed to the new person and setting. If possible, the two members of the evaluation "team" should come from two different professions. This will ensure multidisciplinary as well as multifactored assessment. A speech therapist and school psychologist make an excellent team. Their different approaches will give more information from the same tests.

Parents, especially the primary caretaker, often supply invaluable information. Interviewing the parent first to get a general idea about the child's performance enables the evaluator to begin the testing with items that are close to the child's basal level. With a severely handicapped child, the base performance level may not be immediately evident. By using parent-provided clues, you may help the child achieve some immediate success, and avoid frustration. This information may also give you ideas as to what information assessment measures the child is likely to respond to.

The child should always be allowed to investigate the surroundings. It is then often easier for the examiner and child to begin getting acquainted in that place. Responses that the child makes to his or her environment should be noted. The examiner should note such details as light and sound variations and the child's behavior changes as he familiarizes himself with a new place. These clues may provide important evidence of sensory awareness.

Finally, there are three basic characteristics of the child that should be evaluated early in the session. The results will determine the direction of future evaluation sessions. First, the functional effectiveness of the child's senses will determine the basic types of formal instruments that can be used. Second, the child's ability to communicate should be observed. Finally, evaluation of motor skills such as walking, pointing, grasping, and manipulating objects will provide additional information. Based on this information, the examiner will be able to select the testing devices that are most appropriate and determine the necessary modifications.

Blind/Visually Handicapped Children

Blind or visually handicapped children pose relatively few difficulties for the examiner. Most standardized tests that have verbal subtests may be used with relatively few modifications. One test frequently used for intellectual assessment of the blind is the Hayes-Binet Intelligence Test.

Visual material from the Stanford-Binet can be replaced for the most part by oral and Braille materials. More recently, the verbal subtests of the Wechsler Intelligence Scale for Children--Revised (WISC-R) have been used frequently because they have outstanding standardization.

When assessing blind or visually handicapped children, the examiner might consider asking the child's parents several questions which should assist in determining the extent of the disability. Some questions which might be asked include the following:

-How much useful vision do they think the child has?

-What kinds of behaviors do they observe? For example, how does he respond to sound? to people?

-How does he move around the house--by touch, vision or both?

-How does the child examine toys and other new objects--visually or kinesthetically?

-How does he find objects he drops?

Similar questions can be asked of parents of children with other handicapping conditions. Adaptive behavior scales or developmental profiles are useful as a base for determining a severely handicapped child's functional level.

When evaluating a young visually impaired child, you may want to use larger pictures or models of the stimulus items. Although these materials may not be standardized, you will more likely be assessing his knowledge rather than his ability to see. These materials will give you some idea of the child's functional vision as well as his potential for learning. If non-standardized measures are used, a statement to that effect should be made, as well as descriptions of behavior and performance. Such descriptions will make a report more meaningful and give valuable information for writing an educational plan.

Three informal tests which may be indicators of a visually impaired child's potential to learn are:

1) allowing the child to tactually explore one figure (such as a circle) and select a similar one from a group of figures;
2) Requiring the child to find the one figure that does not belong in a group of figures;
3) have the child put items together by identifying a relationship.

These activities not only enable you to know whether the child has or is able to learn concepts but even more important whether the child is able to relate how objects fit together in specified relationships.

Deaf Children/Hearing Impaired Children

Assessing deaf children requires considerable experience. Many deaf children cannot lip-read or use oral language functionally. As a result, nonverbal tests must be used. Because sight is the chief means by which deaf children receive stimulation, they are likely to use visual cues to confirm their performance. To avoid accidental or differential reinforcement of responses, the examiner should guard against facial expressions and gestures that may provide visual clues to deaf children. Deaf children should not inadvertently be encouraged to make any responses. This issue is critical.

If the deaf child can lip read, appropriate lighting and positioning in relation to the examiner are important considerations. Examiners must make sure that their instructions are understood without indicating the correct answer. Thus all communication and instructions must be geared to the deaf child's level, which may be surprisingly low. Nearly one-third of deaf high school graduates are functionally illiterate, 60% are at grade level 5.3 or below in reading, and only 5% achieve 10th grade reading level or better (Boatner, 1965; McClure, 1966).

Of the special tests devised for deaf, one of the best known is the Hiskey-Nebraska Test of Learning Aptitude, which was standardized in schools for the deaf. At the

preschool level (Sattler, 1974), The Leiter International Performance Scale can be given with pantomime directions. The performance subtests of the WISC-R are also frequently used with deaf children, and they provide some of the most adequate norms for comparison with normal children. Finally, the examiner should keep in mind that performance measures of ability do not predict school performance as well as verbal tests do.

P.L. 94-142 addresses the issue of nondiscriminatory testing and states that test procedures will be conducted in "the primary language" of the child. With children who communicate only through manual systems, this provision could be interpreted to mean that their tests must be conducted via fingerspelling and/or sign language. School psychologists may need to become skilled in fingerspelling (which is relatively easy) and basic sign language (which requires much more time, study, and effort). (See Yarnall, 1979.)

Multihandicapped Children

These children present a demanding challenge for the examiner in that many, if not most, standardized instruments are inappropriate for them. The tests should be administered in a standard manner where possible. When it becomes apparent that the child's ability to respond is so restricted or limited that conventional administration and scoring are

inappropriate, then there are several methods to use to obtain information. First, the examiner can administer only those items which seem to be within the capabilities of the child. Another alternative is to modify the test items so that the child can indicate what he knows by pointing, nodding, or doing any response within his capability (which may be only looking in the direction of the correct answer).

Children who are multiply and/or severely handicapped often have had fewer learning experiences than other children. Therefore, their language--both expressive and receptive--will be limited. This will be especially evident in words which have multiple meanings. For instance the word "rock" may only mean a backward-forward motion, not a large hard substance, or a musical group, or any other of its many meanings. This can always be explored informally by going back over the item and giving it different names, or having the child name it. Further language evaluations should be done in conjunction with a speech therapist who has a strong background in language.

In general, however, where there is a standardized test for assessing an ability of handicapped children, the standard administration procedures should be used. As Cronback (1960) has pointed out, departures from administrative procedures change the meaning of the test scores. If the intent of an evaluation, however, is not to obtain a quantitative index, then departures may be justified.

On those occasions where standard procedures are not followed, that information should be clearly communicated and the resulting scores should be interpreted in light of the test modifications.

Orthopedically Handicapped Children

Depending on the degree of their problem, orthopedically handicapped children may require special procedures. Verbal items will be appropriate for those who have serious limitations. Others who are less seriously involved may need no modifications in administration. Where modifications are required, Brice and Cruickshank (1966) suggest that examiners use tests requiring the least number of modifications. Scores from tests where only some of the items are given may be prorated on the number of items administered.

Children who are physically impaired and usually confined to a wheelchair may be more confortable, and as a result more responsive, while sitting or lying on the floor than while sitting at a table. If the child's impairment is such that he has little or no physical movement, the examiner can do the pointing while the examinee responds by blinking his eyes or tapping his finger or some other response which does not involve large body movement.

If a child performs poorly on verbal items on an intelligence test and yet is fairly verbal in other

situations, you should suspect hearing-related problems. The child may have auditory processing difficulties or a deficit in language due to temporary hearing loss. You might also want to check the child's medical history for chronic upper respiratory problems in order to find out if the child has occasional limited functional hearing. If language deficits are suspected, a comprehensive evaluation by a competent language specialist is necessary.

If a child displays very active or very passive behavior to the extent of not relating at all to the examiner, one approach to training for a response is for the examiner to make the appropriate response with the child and immediately reward him with a primary (dry cereal) and secondary (praise) reinforcer. The caution here is that training can only be done over time and should be done in conjunction with period of observation.

Severely and Profoundly Handicapped Children

Within the past few years, the interest, efforts, and monies that have gone into developing educational programs for children who are severely and profoundly handicapped have greatly increased. Federal funds and school programs for children who are deaf-blind, autistic-like, multiply-handicapped, profoundly disabled, and/or severely functionally retarded are growing rapidly. These children are very difficult to test and teach, and pose special problems

for the school psychologist. At the present time there are no acceptable, standardized tests available to adequately assess most children who are multihandicapped or severely or profoundly handicapped. With these children, school psychologists will need to find tests and subscales that are appropriate for less severely involved children and use whatever portions of trem they can to pull out data that will provide as complete a profile on each child as possible.

Review of 29 Assessment Tools/Tests

ADAPTIVE BEHAVIOR SCALE (ABS)

Purpose: The ABS is designed to assess mentally retarded and emotionally distrubed persons. It provides information on the way an individual maintains personal independence and meets social expectations.

Organization: The scale is applicable to males and females and can be applied to all levels of mental retardation and to all ages, beginning with age 3. Two forms of the scale are provided: one for children 12 and under and an adult form for persons 13 and older. Part 1 of the ABS taps 10 behavior domains: independent functioning, physical development, economic activity, language development, number concepts, time concepts, domestic duties, self-direction, responsibilities, and socialization. Part II provides information on maladaptive behavior such as violent and destructive behavior, untrustworthy behavior, inappropriate interpersonal manners, and hyperactivity.

Administration: The ABS should be administered by informants who are well-acquainted with the daily behavior of the individual to be rated. The scale is relatively easy to administer since the items are written clearly, organized logically, and presented in a well-constructed instruction booklet. Scoring instructions are rather complex; however, instructions are provided for computer scoring which would simplify and reduce scoring time of individual protocols.

Technical Data: The ABS was standardized on a sample consisting of approximately 2300 patients from 63 residential institutions for the mentally retarded throughout the United States. The standardization sample was stratified by sex, intelligence, and age. Although further research on the scale's validity is needed, the face validity appears to be quite good.

Limitations: The standardization sample lacked representatives from younger age levels, minority groups, noninstitutionalized subjects, and upper levels of intelligence. Furthermore, the reliability of the scale has not been determined objectively; no reliability coefficients are provided. Finally, caution should be exercised when using the available

norms because they were derived solely from an institutionalized population.

Considerations: The ABS was intended for use with mentally retarded individuals; however, it can be used with the multiply handicapped population. Since the available norms were derived from an institutionalized population, test users should consider developing norms geared more toward the general handicapped population.

BAYLEY SCALES OF INFANT DEVELOPMENT

Purpose: The Bayley Scales of Infant Development (BSID) are a carefully standardized measure of infant development from age 2 months to 30 months. The primary value of this scale is that it provides a basis for establishing an infant's current status and thus the extent of any deviation from normal expectations. Also, performance on the scales can provide a basis for instituting early corrective measures when a child shows evidence of retarded mental or motor development.

Organization: The BSID consists of three scales: a Mental Scale, a Motor Scale, and an Infant Behavior Record. The Mental Scale consists of 163 items that measure the infant's responses to visual and auditory stimuli, manipulation, and play with objects, social interaction, discrimination of shapes, memory, simple problem solving, and more abstract abilities such as naming objects, understanding prepositions, and number concepts. The motor scale consists of 81 items that measure gross motor abilities such as sitting, standing, walking, and stair climbing,

as well as measuring fine motor coordination. The Infant Behavior Record is a rating scale that measures various aspects of personality such as emotional and social behaviors, activity level responses to objects, sensory areas of interest, and such functions as attention span, persistance, and endurance.

Administration: The instrument is obviously administered to the infant individually. In addition to observing the infant perform the tasks, the examiner is also required to record information about the child's health, prenatal or perinatal difficulties, educational and occupational levels of parents, and family constellation.

The raw scores are changed into index scores by the use of tables in the manual. An index score is a normalized standard score in which each infant is compared to others the same age in the standardized sample. A Mental Development Index is derived from the Mental Scale, and a Psychomotor Development Index is based on the Motor Scale.

Technical Data: The standardization sample included 1262 infants and children, ranging from 2 months to 30 months of age. This sample was representative of the United States population in terms of major geographical areas and residence. Furthermore, the sample was controlled for sex, race, and education of head of household.

Split-half reliability coefficients reported for the various age groups ranged from .81 to .93 with a median coefficient of .88 on the Mental Scale. On the Motor Scale, the reliability coefficient ranged from .68 to .92 with a median coefficient of .84. Those items found to have the highest tester-observer and test-retest reliabilities were those items which focused on eye-hand coordination, sustained attention, memory, and vocabulary.

The validity of the Mental Scale was determined by correlating the scores of the 2-year-olds in the standardization sample with their scores on the Stanford-Binet. Over the 6-months age overlap of the two tests, the correlation was .57.

Limitations: The BSID is difficult to administer because of the large amount of information which must be kept in mind during its administration. Not only does the large number of items in the scale make it difficult, but there are also a variety of factors in the instructions to remember that make these scales difficult to administer.

Considerations: The scale is a useful assessment tool for orthopedically and multiply handicapped youngsters, even above the 2½-year level. Both the Motor Scale and Infant Behavior Record could provide useful information regarding the severity of the handicap. However, for these children, unless they are severely retarded, another intelligence scale would be more appropriate.

THE BENDER GESTALT TEST

Purpose: The Bender can be used to evaluate a child's developmental maturity in visual-motor perception as well as his level of social-emotional adjustment. Koppetz states the Bender is also a useful test for determining a child's ability to succeed in school: "Maturity in visual motor perception is largely a function of age and intelligence ... In school beginners maturity in visual-motor perception is highly correlated with school achievement since a certain degree of perceptual maturity is necessary before a child can learn to read, write, and figure."

Organization: The test consists of a series of nine designs which the child is asked to draw.

Administration: Each design is presented to the child individually. The time taken to complete all 9 designs is recorded and provides the examiner with certain information. For example, if the child takes longer than 7 minutes it could mean that the child is either extremely perfectionistic or that he must work very hard to overcome difficulties in visual-motor

perception. Conversely, if the child completes the drawings in less than four minutes, it could indicate that the child is quite impulsive, immature, or negativistic, or it could show a lack of concentration or effort to carry through the details of the drawing.

If evaluating a child's maturity in visual-motor perception, the drawings are scored according to form, detail of figures, integration of parts, and directionality of designs. However, when evaluating a child's emotional adjustment, emphasis is placed on the size of the drawing, location and organization of the designs on the paper, and the quality of the pencil line.

Technical Data: The Bender Gestalt Test was developed and standardized on public school children. 1106 children were involved in grades kindergarten through fifth grade. The sample included 46 classes in 12 different schools.

Limitations: Data on the test's reliability and validity are minimal. Additionally, due to the nature of the

test, students with a moderate visual impairment are penalized. Thus, the results received would give an indication of their handicap, and not necessarily of their developmental maturity in visual-motor perception. Finally, after age 8, the test does not differentiate at the upper level.

Considerations: The Bender can be fairly well adapted to a hearing impaired population. Additionally, up to the age of 8, it is quite valuable in distinguishing between those children with outstanding visual-motor perceptional and those with immature perception for their respective age levels.

BOEHM TEST OF BASIC CONCEPTS

Purpose: The Boehm Test of Basic Concepts helps isolate individual children whose overall level of concept mastery is low and identify individual concepts with which large numbers of children may be unfamiliar. The author's intent was to provide an instrument to pinpoint deficits in concept mastery, thereby facilitating the remediation of these deficits and promoting school progress.

Organization: The concepts of this test can be grouped into context categories of space, quantity, time, and miscellaneous. To a certain extent, the initial item content was empirically and subjectively determined by inspection of numerous curriculum materials, plus arbitrary probes to see what concepts were difficult or unfamiliar to a substantial number of children.

Administration: The Boehm Test of Basic Concepts is appropriate for use with children in preschool through second grade. Testing time is usually 15 to 20 minutes for each of the four booklets; however, this may be longer when testing blind children.

Instructions state that groups of 8 to 12 children may be tested at one time; however, when assessing blind children, it may be advisable to administer the test individually. Scoring instructions are clear, and the mechanics are very simple.

Technical Data: The standardization sample consisted of 9500 children from 16 cities located across the United States. There was an attempt to include all socioeconomic levels in the sample.

Limitations: No validity evidence other than face validity has been presented.

Considerations: Persons familiar with kindergarten and first grade instruction will realize that the test items tap concepts that children need to know. Additionally, the author presents the Boehm as a screening device and as a guide for teaching activities. When used in this manner, the test can provide the examiner with valuable clues regarding basic language and concept comprehension of auditory stimuli presented to young children in sentence form rather than single-word form.

CALLIER-AZUSA SCALE

Purpose: This instrument is a scale of normal development designed to be used in assessing deaf-blind and multihandicapped children. The scale is based on the assumption that, given appropriate educational and therapeutic opportunities, all children follow the normal developmental sequence. It can be used by teachers and psychologists for the following three purposes:

1) To assess the developmental levels of children for whom the traditional methods of assessment provide insufficient information (i.e., due to dependence on language, lack of cooperation, or deficits of either the auditory or visual modality).
2) To provide guidance for planning developmentally appropriate curriculum programs for individual children.
3) To provide progress data over a period of time or following a specific intervention procedure to help measure the overall effectiveness of a specific educational or therapeutic technique.

Organization: The Callier-Azusa Scale is currently comprised of four areas: motor development, percentual development, daily living skills, and language development. The authors plan to add a fifth scale, socialization, at a future date. Each area has three or four subscales made up of six to 17 developmentally sequential steps describing developmental milestones for children.

Administration: The scale must be administered by individuals who are thoroughly familiar with the child's behavior No specific expertise is required other than good observational skills and a knowledge of the child's behavioral repertoire. Most valid results are obtained if several individuals having close contact with the child rate him.

Technical Data: None.

Limitations: More comprehensive information on the validity of the scale and the reliability of the items needs to be provided. Furthermore, more precise information concerning standardization procedures is required.

Considerations: The Callier-Azusa Scale represents an excellent attempt to assess the development of multihandicapped children in specific terms. This instrument shows great promise as a helpful tool in both clinical and educational settings. However, recognizing that the scale depicts the child's progress only, and not his or her potential, it should be viewed and administered only in conjunction with other assessment instruments.

THE COLUMBIA MENTAL MATURITY SCALE

Purpose: The Columbia Mental Maturity Scale (CMMS) is a nonverbal test designed to yield an estimate of the general mental maturity of young children. Work was begun on the scale in 1947 in an effort to provide a more satisfactory instrument for estimating the mental capacity of children with cerebral palsy or other handicaps involving motor or verbal functioning. The authors do not claim that their scale measures the innate capacity of the child; rather it "measures general reasoning abilities which are especially important for success in school, where the ability to discern relationships among various types of symbols is emphasized."

Organization: The CMMS consists of 92 pictorial and figure classification items, each printed on a card 6 x 9 inches. These items are arranged in a series of eight overlapping scales or levels. The authors feel that the objects depicted on the cards are, in general, within the range of experiences of most American children, even those with environmental deprivation. The items range from the

perception of rather gross differences in color or form to recognition of very subtle relations in pairs of pictures so as to exclude one. Except for a group of perceptual items at the beginning of the first three levels, no two successive items are based on the same concept or employ the same rule.

Instead of an IQ score, the CMMS gives an age deviation score (ADS). This indicates the extent to which the child's performance deviates from the average performance of children of his same chronological age.

Technical Data: The 1970 standardization was designed to yield a stratigied quota sample of 2600 children aged 3-6 to 9-11 years that was representative of the total population of the United States. The scale was broken into 6-month age groups with 200 children in each level. The sample contained 50% boys and 50% girls. The United States was divided into four geographical regions, and the samples were representative of those populations. Within each of the regional samples, there was a further

division based on racial and parental occupation groups.

The authors used the Spearman-Brown Prophecy Formula to determine the split-half reliability. The median split-half reliability coefficient across all levels approached .90. The median test-retest reliability coefficient for age groups 6-0 to 8-5 was .85, with an average gain of 4.6 points in ADS on retest.

The CMMS was compared with various subtests of the 1964 Stanford Achievement Test for students grades 1 and 2 in a Virginia school district. The correlations had median value of .57 for all Primary I Battery subtest areas. At the second grade the correlations had the median value of .47 for all Primary II Battery subtest areas. Using the same population, the CMMS correlated .69 with the Otis-Lennon Mental Ability Tests for pupils grades 1 through 3. When compared to a preschool and first grade population, the CMMS correlated .84 with the Stanford-Binet Intelligence Scale, From L-M. Additional validity studies are currently being undertaken.

Limitations: A major limitation of the scale is that there are no specific norms for those handicapped populations for whom the authors feel that the test is ideal. Therefore, caution is needed when interpreting the scores of a physically handicapped individual and comparing them with norms which have been generated in normal children.

Considerations: The CMMS is especially suitable for use with children who have cerebral palsy (or other brain damage), mental retardation, speech impairment, or hearing loss. Since the scale does not depend upon reading or other language skills, it is useful with children coming from homes where English is not the primary language. Finally, the manual includes Spanish directions, and an interpreter could be used when administering the scale in other languages.

DENVER DEVELOPMENTAL SCREENING TEST

Purpose: The Denver Screening Test was designed to be used by people with no training in psychological testing for the purpose of detecting children with serious developmental delays.

Organization: The Denver evaluates four aspects of a young child's functioning: gros motor, fine motor adaptic, language, and personal-social development. It consists of 105 potentially discriminatory items selected from 12 existing developmental and pre-school tests.

Administration: The scale is administered by informally or formally observing the child under consideration. Each of the four areas are scored as either normal, abnormal, or questionable. A child should be considered abnormal if two or more delays are noted. Finally, a child's level of functioning in a particular area should be considered questionable if one delay is recorded.

Technical Data: The Denver was normed on 1036 children, 543 males and 493 females, between the ages of 2 weeks and 6.4 years. Children known to be

adopted, premature, or handicapped were excluded from the standardization sample.

Research on the scale's reliability has generated coefficients between .66 and .93. These coefficients were generally higher when calculated on children ages 3 and 4 as opposed to ages 1 and 2. Also, although little information is available on the scale's validity, the Denver has been found to correlate with the Yale Developmental Examination at .97.

Limitations: Norming for the scale is weak. Supposedly the sample closely approximated the ethnic and occupational groups included in Denver's 1960 census. However, the sample contained a significantly higher number of white children and children whose fathers functioned in professional or managerial positions than the census distribution would warrant. Also, the majority of the items are based on actual observation of child behavior and have only moderate reliability. Finally, there is little data on the validity of the scale, and the reliability of the items designed for the younger age levels is questionable.

Considerations: The test is useful with the orthopedically handicapped for the purpose of determining their developmental level of functioning. It can also be modified for use with the hearing and visually impaired as a quick screening device to determine which assessment scales should be employed to assess a specific child. Furthermore, the Denver can be used to make a rough estimate of the developmental levels of the child in comparison to normal children. Finally, it is important to bear in mind that the Denver was standardized on a normal population; thus a handicapped child may obtain a spuriously low developmental quotient.

GESELL DEVELOPMENTAL SCHEDULES

Purpose: The Gesell Developmental Schedules were created to assist in determining the extent to which infants have acquired specific behaviors in key developmental areas.

Organization: The Schedules evolved from a longitudinal study of infant development in a sample of human infants. The focus is on four major areas: motor behavior, adaptive behavior, language behavior, and personal-social behavior. In general, the Gesell Schedules represent a standardized procedure for observing and evaluating the course of behavior development in a child's daily life.

Administration: Although a few of the schedules may be properly described as tests, most are purely observational. A set of standard toys and other test objects is employed in making these observations. The direct examination of the child is supplemented by interview data obtained from the parent.

There is no scoring system for the test. Instead, the record indicates the appropriate developmental level, in months, that the child

has attained in each of the four areas covered. This is found by comparing the child's behavior with that given as typical of eight 'key ages' between 4 weeks and 36 months.

Technical Data: The standardization population included a small sample of nonhandicapped infants. To date, no statistical analysis of reliability or validity has bee reported.

Limitations: Both observation and scoring procedures are less highly standardized in the Gesell Schedules than in the usual psychological test. Furthermore, a certain amount of subjectivity enters into the examination at several points. The restriction in both size and nature of the normative sample should also be kept in mind. This consideration is particularly important when using the schedules with handicapped children. Finally, technical data are extremely limited.

Considerations: In general, the Gesell may be regarded as a refinement and elaboration of the qualitative observations routinely made by pediatricians and other specialists in infant development.

GOODENOUGH-HARRIS DRAW-A-MAN TEST

Purpose: The Draw-A-Man Test is frequently used as an initial screening tool to identify those individuals who will need a more comprehensive evaluation. Since most children enjoy taking this test, it is often used as an initial test in a battery for the purpose of establishing rapport.

Organization: This test involves the child making three drawings, a man, a woman, and a drawing of self. It focuses on the accuracy of an individual's observation and his development of concepts rather than on artistic skills. It is designed so that almost any adult can learn to score the drawings with reasonable accuracy. Harris stresses, however, that training in psychological testing is needed to adequately understand the test results.

Administration: The Draw-A-Man is quickly administered. Drawings are scored for 73 items (71 items for the woman drawing), which were selected because of age differentiation, relations to total score, and their relation to group intelligence scores. The test uses standard scores that express a child's standing on the test in relation to his own age and sex

group. Since the drawings of self were not standardized and no norms are included for them, Harris recommends averaging the standard scores of both the man and the woman drawings to obtain a more reliable estimate of the child's maturity. The author also includes 12 example drawings or Quality Scales. This is a more rapid method of scoring. The examiner compares a drawing with one of the 12 examples. The Quality Scales are useful when only a rough estimate is needed of a child's maturity.

Technical Data: This test was standardized on 2975 children ages 5 to 15 years. The sample was considered to be representative of the occupational distribution of the United States in 1950 in 4 major geographical areas. This scale generally showed interscorer reliabilities over .90. Split-half reliabilities found on the earlier version of the scale were in the .70s and .80s. The retest scale ranged from .60 to .70. The inner scorer reliability of the Quality Scale fell within the .80 level of significance. Finally, validity correlations with the Stanford-Binet ranged from .36 to .65.

Limitations: The Draw-A-Man Test is only useful as a crude index of mental development when no appropriate standardized tests are available. It should never be used alone to determine if a child needs to be placed in a special class.

Considerations: This test can be used with deaf children by fingerspelling the directions or writing them down on paper. It would even be possible for a deaf child to indicate what certain items in the drawing were and to pantomime what the person in his drawing was doing. The test is extremely useful in roughly estimating a child's abilities. Finally, it is very useful in indicating patterns of ability or in making decisions when there are discrepancies between other test scores.

HAPTIC INTELLIGENCE SCALE

Purpose: The Haptic Intelligence Scale was developed as a nonverbal test to measure the intelligence of blind adults. It is one of the few instruments that can be used with the blind. Additionally, the scale evaluates several facets of intelligence that are not adequately assessed by many present verbal tests.

Organization: Four of the five subtests of the original scales were modified adaptations of the Wechsler Intelligence Scale (WAIS). Wechsler's procedures were followed with respect to establishing age categories and treating statistical data.

Administration: Psychological examiners who are trained in the administration and interpretation of the WAIS should find the Haptic scale relatively easy to administer. Since the administration time is rather long, the examiner should plan on administering the test in two sessions. Finally, it is recommended that psychologists administer this scale in conjunction with verbal tests of intelligence in order to provide a more comprehensive evaluation of blind persons.

Technical Data: The standardization sample was generated from four designated geographical regions within the United States. Visual acuity did not exceed 5/200 in the better eye with correction. As with the WAIS, the distribution of scores was converted to a scale, with a mean score of 10 and a standard deviation of 3.

Limitations: There are several limitations in using this instrument with partially sighted subjects. First, since the test is based on a tactile approach and even the slightest view of the materials would invalidate the subject's performance, the partially sighted person must wear a blindfold. This often becomes uncomfortable before the lengthy session is over. Second, it may be questionable whether partially sighted persons have acquired the highly developed tactile skills which characterize the blind subjects used to establish the test norms.

Considerations: Authors of the Haptic Scale recognize that the instrument requires further refinement and state that it should be administered and interpreted with caution. Furthermore, they stress that, while

the obtained results correlate well with the WAIS, the implication should not be made that the Haptic Intelligence Scale measures the same factors that the WAIS Performance Scale assesses in sighted individuals. However, considering all factors, the Haptic Scale does appear to be a clinically useful tool that provides valuable clues to those who work closely with blind adults.

HISKEY-NEBRASKA TEST OF LEARNING APTITUDE

Purpose: The Nebraska Test of Learning Aptitude for Young Deaf Children was the first available test of ability which has been specifically designed for young deaf children and standardized upon such children in this country. The Hiskey-Nebraska Test of Learning Aptitude (H-NTLA) represents a revision and restandardization of this original scale.

Organization: The H-NTLA is a non-verbal test which was normed on both a deaf and hearing population. There are 163 items and 12 subtests; bead patterns, memory for color, picture associating, paper folding, visual attention span, block patterns, completion of drawing, memory for digits, puzzle blocks, picture analogies and spatial reasoning.

Administration: The scale was designed for use with children ages 3 to 17 and takes about 40 to 45 minutes to administer. Not all subtests are administered to all ages, some are designed for ages 3 to 10, others are designed for use with ages 11 and above, while still others are used for all ages.

The scale makes use of age norms. Since the H-NTLA compares deaf children with other deaf

children, the age norms are referred to as Learning Age (LA) and Learning Quotient (LQ) as opposed to MA and IQ. The LQ is found by dividing the LA by the chronological age.

Technical Data: The H-NTLA was standardized on children from ages 2-6 to 17-5 years of age. The population consisted of 1079 deaf and 1074 hearing children from 10 different states. The samples corresponded closely to the United States Census data although no breakdowns were given by race or sex. There was parental occupational sampling for the hearing populations but none for the deaf. The majority of the deaf population came from the state schools for the deaf.

Using the split-half method and the Spearman-Brown formula, the coefficients of reliability were: .95 for deaf children ages 3 to 10, .92 for deaf children ages 11 to 17, .93 for hearing children ages 3 to 10 and .90 for hearing children ages 11 to 17.

Concurrent validity was possible only with the hearing population. The correlations with the Stanford-Binet ranged from .78 for the older children to .86 for the younger children. Correlation with the WISC was .82. For the deaf population subtest intercorrelations ranged from .33 to .74 for the age range of 3 to 10 and from

.31 to .43 for the age range of 11 to 17. For the hearing population subtest intercorrelations ranged from .32 to .78 for the age range of 3 to 10 and from .25 to .46 for the age range of 11 to 17.

Limitations: Currently, opinions appear to vary relative to the usefulness of this scale. Some individuals seem to prefer the H-NTLA over the Leiter while others state that the H-NTLA requires too much manipulation and is more cumbersome to administer.

Considerations: Burros describes the H-NTLA as most promising for the assessment of the "book learning" capability of the deaf. He indicates that the subtests: memory, picture identification and picture association are the most promising for the assessment of the deaf. Furthermore, a distinct advantage of this instrument is that it includes norms for both a deaf and a hearing population. Directions for both the deaf and the very hard of hearing are given through pantomime. However, if a hard of hearing child is able to receive the verbal directions, he is then compared to the hearing and not the deaf norms.

INTERIM HAYES-BINET

Purpose: The Hayes-Binet was developed to measure intellectual ability and academic potential in blind children. It combines verbal forms L and M of the Stanford Binet and was used widely prior to the publication of the Wechsler scale.

Organization: The test covers from age 3 years to Superior Adult III. The majority of the items have been taken directly from the two forms of the Stanford-Binet, although several of the items have been written specifically for this test. Braille material is provided for most, but not all, of the subtests. The author felt that the few nonbraille material items would not adversely affect the test performance if presented orally.

Administration: The test is designed for administration to young children (age 3) through senior adults. Administration protocol follows essentially the same procedures established for the Stanford-Binet.

Technical Data: Validity studies comparing the Wechsler scale and the Hayes-Binet indicate that they cannot

be considered interchangeable with blind children of above average ability. For blind children with average or below average ability, the scales appear to yield more comparable IQs. Also, the Hayes-Binet has been found to be somewhat more valid than the Wechsler scales when teacher's ratings are used as a criterion.

Limitations: Considering current advances in testing, much of the content of the Hayes-Binet is very outdated and awkward. Consequently, psychologists presently working with blind individuals are more likely to rely on the verbal portion of the Wechsler scales.

Considerations: In recent years, a psychologist at the Perkins School for the Blind in Massachusetts has worked on and updated recent versions of the Binet; it has been called the Perkins-Binet. This test is predominantly a verbal scale, but performance items have been added to achieve better balance. When the standardization of this instrument is completed, it should be a better tool for evaluating the learning ability of blind students.

THE LEITER INTERNATIONAL PERFORMANCE SCALE

Purpose: The Leiter International Performance Scale (LIPS) is a nonverbal test of intelligence that can be used to evaluate children with sensory, motor, speaking, or reading deficits. The scale has been designed to measure intelligence by measuring an individual's ability to analyze and synthesize materials.

Organization: In principle, the Leiter is a nonverbal Stanford-Binet Scale containing 54 tests arranged in an age-scale format. The scale appears to require perceptual organization and discrimination ability. The tests at the lower levels of the scale are considered by Sattler (1974) to be tests of the ability to learn rather than of tests of learned materials.

Administration: The LIPS is designed for administration to children ages 6 through 17. Each task requires that the child first select those blocks containing appropriate symbols or pictures. The child then inserts these blocks into the correct notch on the frame that is used to administer the scale. Except for the subtest block design, the LIPS is

is an untimed test. Instructions are given in pantomime. When evaluating hearing subjects, it is important to explain to them that since this test was designed for use with the deaf, they cannot be given any explanations. Leiter suggests that in order to eliminate questions that the subject might ask, the examiner should "look busy" while the subject is completing the task.

Technical Data: The only information provided regarding the norming of the LIPS is that the 1940 scale was standardized on a Hawaiian population consisting of Chinese and Japanese. When administered to children in the continental United States, the scores obtained were approximately 18 months above those compiled on children in the standardization sample. Recognizing this, Leiter then modified the norms (although he did not specify how this was done), thus making scores spuriously low by 6 months. Consequently, IQ and mental age adjustments are required if valid final scores are to be reported.

Leiter does not report any information regarding the validity of his scale except to say that the 1948 revision correlates at .92 with the 1940 scale. However, Sattler (1974) states that the LIPS correlates with the Stanford-Binet and the Wechsler Intelligence Scale for Children in a range of .56 to .92 with a median correlation of .83. As would be expected, the LIPS correlates higher with the performance scale of the WISC than with the verbal scale.

The manual includes no information on the reliability of the scale. Once again, Sattler (1974) reports studies that show a test-retest coefficient of .91 with mentally retarded children and of .86 with language-impaired children.

Limitations: Because of the way that the scale was standardized, it does not give IQ's and MA's that are directly comparable to other scales. Therefore, a considerable abount of time is required to calculate and find adjusted IQ's and MA's. Two additional disadvantages are that the tables in the back of the General Instructions are inappropriate for

certain age groups and that the tables for determining the adjusted MA and the instructions for administering and scoring the ITEMS are in two separate manuals. Finally, several additional limitations have been attributed to the LIPS. The item difficulty levels may be uneven; certain pictures are outdated; the culture fairness of the scale has not been determined; the abilities measured by the scale are not clear; and the scale contains a small number of tests at each year level.

Considerations: Although limitations relative to the construction, utilization, and interpretation of the Leiter are apparent, the nonverbal format of the scale makes it ideal for use with hearing impaired populations. It is also useful for testing handicapped children who could not otherwise be evaluated by the Stanford-Binet or the WISC. Additionally, it is useful in assessing those children who might be penalized by picture vocabulary tests, such as certain ethnic groups or culturally deprived groups.

The Arthur Adaptation of the Leiter Performance Scale

Dr. Arthur adapted the 1948 revision of the LIPS. She supplemented this revision with tests from the Revised Form II of the Point Scale of Performance Tests. For the most part, the changes in the LIPS are of a rather minor nature. Since Arthur felt that the norms on the Leiter for ages 3 to 8 were too high, her adaptation covers these ages and includes test levels II through XII. The test materials are exactly the same as the Leiter materials up to level XII, where her adaptation ends.

Although the materials used in the two scales are identical, there is a slight difference in administering the scales. On the subtest Concealed Cubes, level X-3, Arthur's adaptation includes a demonstration of what is to be done by the subject. This subtest involves the use of hidden cubes, and Leiter suggests that the value of the test lies in letting the subject discover that there are hidden cubes.

Other changes in administering the Arthur Adaptation are so minor that at first the Psychological Corporation would not permit a separate printing of Arthur's adaptation, and her changes were included in italics in Leiter's manual. There is no separate manual for the Arthur Adaptation.

MAXFIELD-BUCHHOLZ SOCIAL MATURITY SCALE FOR BLIND PRESCHOOL CHILDREN

Purpose: The Maxfield-Buchholz Social Maturity Scale is an adaptation of the Vineland Social Maturity Scale designed for use with young blind children, who generally respond poorly to formal tests.

Organization: The scale consists of 95 total items, each placed within a given year level of expected performance. The items are grouped into the following seven categories: self-help general, self-help dressing, self-help eating, communication, socialization, locomotion, and occupation.

Administration: The scale covers the chronological age range of birth to 6 years. It is administered using an interview approach with an informant having intimate knowledge of the child. Validating evidence is gained from direct observation of the child by the examiner.

Technical Data: The scale was normed on 484 children in the United States using the legal definition of blindness as a basis for determining the degree of vision. Although the scale's sample was small, there was

a diversity of social backgrounds, heredity, developmental opportunities, and eye conditions.

Limitations: A true measure of reliability for the scale has not yet been developed. Also, while the test covers the entire preschool range, the norming sample included only those children whose ages fall between 5 months and 5 years, 11 months. A further disadvantage is that results are based on the child's reported typical behaviors. This frequently creates difficulty in that responses to interview tests are easily affected by the child's or parent's mood, health, and energy on the day of the exam. Furthermore, fright or fatigue, plus any number of other factors, may affect the test results. Therefore, this scale should be interpreted cautiously and in conjunction with other available assessment tools.

Considerations: This scale can be used most effectively by an examiner who has considerable experience with the diagnostic interviewing process with both parents and children and who has experience in the psychological testing of small children. A major

advantage of the scale is that it can be used as a guide for counseling the parents and teacher toward improved skills.

MICHIGAN MANUAL FOR THE ASSESSMENT OF DEAF-BLIND, MULTIPLY HANDICAPPED CHILDREN

Purpose: This scale is designed to provide the teacher of young deaf-blind and multiply handicapped children with an assessment tool to use in setting goals for the various areas of development. According to the authors, this instrument should be used primarily to establish goals for each child and to subsequently evaluate his progress at specific intervals.

Organization: The scale was developed by the Deaf-Blind Department at the Michigan School for the Blind. It was adapted from existing scales, primarily the Developmental Scales for Deaf-Blind Children from Denver, Colorado. Additionally, since a large number of severely handicapped children have severe developmental deficiencies, special emphasis is placed on the initial steps of various developmental sequences. Finally, recognizing that development does not occur in perfectly sequenced steps for all children (i.e., steps may be smaller at one point and larger at another), the authors

attempted to develop a technique for comparing levels across scales. However, they were unable to accomplish this due to the variation among the deaf-blind population and because scales were not set up according to mental levels.

Administration: This test is designed for administration to deaf-blind children who have concurrent problems in communication development and education.

Limitations: The Michigan Manual is still rather rough and therefore needs further refinement. Also, more comprehensive information on the validity of the scale and the reliability of the items needs to be provided. Finally, more precise information concerning standardization procedures is required.

Technical Data: None.

Considerations: In its present state the Michigan Manual is a good supplementary instrument for this population. The most important function it serves is to measure change in a child's growth over a period of time. This purpose is facilitated by planning sheets that provide sufficient space for repeated re-writing of functions and goals.

ONTARIO SCHOOL ABILITY EXAMINATION

Purpose: The Ontario School Ability Examination is a non-verbal performance test designed to measure a subject's current intelligence. The test evaluates intelligence by assessing the subject's ability to eliminate trial-and-error behavior, by taking into consideration the facts of a situation and planning a course of action adequate to the occasion.

Organization: The scale is composed of modifications of the Gesell Block Building and Gesell Paper Folding Sequence, the Drever Collins Block Design, the Drever Collins Domino, the Knox Cube, the Healy Fernald Puzzle, the Tapping Examination, and certain items from the Stanford Revision Examination. The instrument is divided into subtests, series, and mental levels. The six areas are manipulation, color patterns, knox blocks, dominoes, drawing, and tapping.

Administration: The Scale is designed primarily for use with children ages 3 to 15. It uses a mental age and a ratio IQ. If the subject is over 15 years

of age, the maximum chronological age division would be 180 months. A mental age of over 15 or under 3 should be considered suggestive only.

Technical Data: The Ontario Examination was normed on 288 students at the Ontario School for the Deaf, ranging in age from 5 to 22 years. Amos states that his test is "valid in the Province of Ontario and probably valid in other comparable areas for the purpose of determining the school ability up to and including mental age 12 years of deaf children, retarded children, and children whose home language is other than English.

Limitations: The instrument lacks data on the reliability and validity of the scale. Furthermore, since the author provides a poor description of his strikingly samll standardization population, caution is needed when interpreting scores. Recognizing these limitations, the scale can never be used as a sole indicator of ability.

Considerations: The Ontario School Ability Examination has been described as a reasonably good test for deaf children within the age range of 4 to 10 years.

Additionally, the nonverbal nature of the test not only makes it useful in assessing deaf children, but also enhances its use with certain ethnic and culturally deprived groups.

PEABODY INDIVIDUAL ACHIEVEMENT TEST

Purpose: The Peabody Individual Achievement Test (PIAT) was designed to estimate an individual's current level of academic achievement. Where an intelligence test generates information about an individual's academic potential, achievement tests such as the PIAT provide the examiner with a general breakdown of current academic skills.

Organization: The PIAT consists of subtests in the areas of math, reading recognition, reading comprehension, spelling, and general information. The items within each subtest range sequentially from kindergarten to high school. The PIAT provides four types of derived scores. Two are developmental (central tendency) scores in grade and age equivalents. The authors define grade equivalents as the grade level at which that score represents the median and age equivalents as the chronological level at which that score represents the median. The other two deviation scores are percentile ranks, both within grade and age level, and standard scores.

Administration: The PIAT was designed as an untimed screening test which can be quickly administered within 30

to 40 minutes. Only those items falling within the subject's range are administered. This range is determined by grade placement in math, and the raw score of the preceding subtest on the other subtests. On all subtests except general information, the subject is shown a response card with four possible answers. The subject points to or says the number of what he feels is the correct answer. On the general information subtest, there are no response cards.

Technical Data: The PIAT was standardized on a national population of children in regular classrooms in public school systems. It was normed on 2889 children, 1401 males and 1488 females. The population consisted of 2406 white children, 321 black children, and 123 additional minorities, as determined by the ratio of minorities in the 1967 census.

In terms of median values, the greatest confidence can be placed on the stability of the total test score (.89) and on the reading recognition score (.89). The least confidence can be placed in the reading comprehension score (.64) and the

spelling score (.65). Product-moment correlation coefficient between subtests vary considerably, (.22) between math and reading comprehension and total test at the third-grade level. The standard error of measurement is smaller at the lower grades, where the test is more sensitive, than at the upper grades.

Concurrent validity with the Wide Range Achievement Test is higher than any other individual achievement test, math (.58), reading recognition (.95), and spelling (.85).

Limitations: When compared with group achievement tests, the PIAT has less demonstrated validity, lower test-retest reliability, and a smaller standardization sample. Burros also states that more research is needed before the PIAT can be accepted as a valid test.

Considerations: Since the PIAT is administered without any writing required, it lends itself well to use with orthopedically handicapped children. Instead of pointing to the correct answer, the child can verbally indicate his choice. Another method

could be for the examiner to point to each answer and the child indicate through some signal (wink, nod, smile, etc.) his choice of correct answer. It could also be adapted for use with hearing impaired children. Instead of reading the instructions and questions, they could be written for those children with good reading ability, or the examiner could use signing or finger spelling. However, caution is needed since even the hearing impaired child who can read may lack many of the verbal concepts known by hearing children of the same age. It may be necessary to determine the child's understanding of verbal concept through the use of a vocabulary test (i.e., Peabody Picture Vocabulary Test). Finally, caution is also needed when interpreting the results of the test, since the PIAT was normed on a completely normal population that did not include handicapped subjects.

PEABODY PICTURE VOCABULARY TEST

Purpose: The Peabody Picture Vocabulary Test is a nonverbal test designed to provide an estimate of an individual's verbal ability through measuring his hearing vocabulary or his receptive knowledge of vocabulary.

Organization: The Peabody consists of 150 plates with four pictures on each plate. These plates are arranged in an increasing level of difficulty from 1-9 to 18 years. There are two forms of the Peabody, Form A and B, which vary in that they use different words. Items are presented in a multiple-choice format. The manual provides percentages and mental ages for each raw score.

Administration: The test is administered to children who are 2½ through 18 years of age. Testing time is between 10 and 15 minutes. Only those items that fall within the critical range for each student are administered.

Technical Data: Correlations between the Peabody and Stanford-Binet range from .22 to .92 with a median correlation of .66. Studies comparing the Peabody and the Wechsler Intelligence Scale for Children indicate

that the biggest correlations are obtained with the verbal scale; however, correlations between the Peabody and the Performance Scale are considerably lower.

The standardization samples included 4012 white subjects (ages 2½ to 18 years) residing in and around the Nashville, Tennessee, area.

Limitations: Scores obtained on the Peabody must never be exchanged with those obtained from a more comprehensive test of verbal ability. Additionally, special care must be given in examining ethnic minority children since they frequently obtain lower scores on the Peabody than on other intelligence measures. Therefore, although the Peabody is a useful tool, it must be interpreted cautiously when evaluating the intellectual ability of exceptional and minority group children.

Considerations: The Peabody has wide use as a clinical tool. In addition to being effective with average students, it has special value when employed with special needs groups. Since subjects are not required to read, the scale is especially fair to nonreaders. Recognizing that responses are nonoral, the scale

is appropriate for the speech impaired. Furthermore, since neither a pointing nor an oral response is necessary, cerebral palsied and other orthopedically handicapped persons are not penalized. It is possible to enlarge the Peabody plates for use with the partially sighted, and it has been found that children with severe visual handicaps obtain higher scores with enlarged pictures than with standard pictures. For hearing impaired children, it is also possible to present the Peabody by typing the words on cards or finger spelling them rather than presenting them orally. Finally, the Peabody might be useful as a device to estimate the acquired verbal concepts of the hearing impaired child before administering another scale that requires the child to read questions and answers.

PICTORIAL TEST OF INTELLIGENCE

Purpose: The Pictorial Test of Intelligence is an easily administered, objectively scored, untimed, individual test of intelligence. It has been designed for use with both normal and handicapped children between the ages of three and eight.

Organization: Except for the use of a multiple-choice format, the test is in some respects a further development of Binet type scales. It presents tasks which are designed to measure verbal comprehension, perceptual organization, and the ability to manipulate increasingly complex symbols (spatial and numerical). The tasks are arranged into six subtests: Picture Vocabulary, Form Discrimination, Information, Comprehension, Similarities, Size and Number, and Immediate Recall.

Administration: The scale is administered individually to children ages three through eight. The examiner does not find a ceiling or a basal, all items are administered. The questions are presented orally by the examiner while he shows the child some large picture cards on which are printed four possible choices. The total correct items are converted by a table to MA units and then to a deviation IQ.

Technical Data: The Pictorial Test of Intelligence (PTI) was standardized in 1962 on a total sample of 1830 children, ages three to eight. At each level, the samples ranged in size from 260 to 347 children. The samples were drawn from the Northeastern, Northwestern, Southern, and Western regions of the United States. The sample was further stratified by community size and by the father's occupational level to be representative of the 1960 Census. Both white and non-white children were included in the sample.

Test-retest reliability ranged from .90 to .96 for intervals of two to six weeks. The Kuder-Richardson reliabilities ranged from .87 to .93. The lowest estimates of reliability were at the upper and lower age limits of the test.

Concurrent validity was established by correlating the PTI with the Stanford Binet, .72; the WISC, .65; and the Columbia Mental Maturity Scale, .53; with first grade children. Sattler (1974) reports of studies using a sample of 7 to 14 year old mentally retarded children. The PTI correlated at .77 with the Peabody Picture Vocabulary Test. Using the same population, the predictive validity was determined by comparing it with the spelling (.68), reading (.56), and arithmetic (.79) of the Wide Range Achievement Test.

Limitations: Caution is needed when using written instructions with the hearing impaired person so that he is not penalized due to inadequate reading skills.

Considerations: In order to respond to the test items, children who understand English need only be capable of hearing simple verbal instructions and responding to visual stimulation. The child can indicate his response to the question by pointing to the pictorial symbol in the card. Additionally, the choice of responses on the response cards are designed so that eye movements, such as staring at the chosen response, can be used for those children who are orthopedically handicapped and cannot point. Finally, since the verbal instructions are rather simple, it would be possible to adapt the scale for use with the hearing impaired by utilizing finger spelling or written instructions.

PORTEUS MAZE TEST

Purpose: The Porteus Maze Test was originally designed to provide a more accurate diagnosis than could be obtained by other intelligence scales. It is a performance test of intelligence, foresight, and planning, for use with individuals over 3 years of age.

Organization: The test involves three different maze series. The original or Vineland is always given first. The Extension Series and the Supplement Series are intended to be retest measures. The original series consists of 12 mazes; the other two series each have 8 mazes.

Administration: Porteus suggests that there are certain behaviors needed to go through a maze. First, the subject must recognize a goal. He then must identify subgoals. Shortterm memory for the preferred course of action is needed. And finally, the subject must carry out his plan.

There are three ways of scoring the Maze Test: age units, Q scores, and a conformity-variability score. Porteus observed that the test-taking

styles were different among various youngsters even when their age scores held constant. He observed a difference in impulsiveness, carelessness in performance, and failure to follow directions. Various penalty scores are assigned to a subject's test-taking behavior regardless of his success in dealing with the mazes. This gives a qualitative or Q score. There are nine test performance characteristics (e.g., cutting corners, crossing lines, lifting pencil) which are assigned penalty scores.

Technical Data: Burros states that this test has limited standardization. He further specifies a correlation between the original series and the extension series at .50. The reliability of the Q score has been calculated above .95.

Burros describes the Porteus Maze Test as having "considerable face validity as a task requiring short-term planning toward the solution of printed mazes." Its correlates with various general intelligence tests range from .40 to .70. It correlates with Knot Cubes at .75. The Q score correlates with social

adjustment; it differentiates delinquent youth from nondelinquent youth. The correlations between Q scores and various IQ scores range from -.20 to -.40.

Limitations: Porteus wrote a book rather than a manual for this test. This text does not meet the standards set by Standards for Educational and Psychological Tests and Manuals. Also, Porteus' book contains no information on reliability and only a poor description of the norms.

Considerations: Since this is a nonverbal performance test that is possible to administer without verbal instructions, it can be used with the hearing impaired. However, the correlations range from .40 to .70; consequently, when used for this purpose, it should be used with caution and then only as second source of information. It is not recommended as a sole indicator of mental ability. Finally, the Maze Test seems most useful in the identification of the adaptability of poorly acculturated persons.

PRESCHOOL ATTAINMENT RECORD (PAR)

Purpose: The PAR is designed to measure the physical, social, mental and language attainments of children from 6 months to 7 years old. This instrument is especially useful in cases of sensory impairments, speech or language difficulties, emotional disturbances, neuromuscular impairments, resistance to examination, and cultural differences. More specifically, it is well adapted for testing deaf, blind, and multiply handicapped children, children with orthopedic disabilities, and children whose language development is delayed.

Organization: The scale is essentially an expansion of the early age levels of the Vineland Social Maturity Scale. Like the Vineland, the purpose of the PAR is to determine the child's usual behavior in the areas tested. The instrument is divided into eight areas, with 14 items in each category arranged in order of increasing difficulty. The eight areas are: ambulation, manipulation, rapport, communication, responsibility, information, ideation, and creativity. The scale permits comparisons of child

with child, or of a given child with himself, in successive measures and is useful for composing homogeneous groups or assessing their comparability.

Administration: PAR is administered as a structured interview with an adult who is thoroughly familiar with the child and/or by direct observation of the child. Items are scored +, -, ± for pass, fail, and doubtful. An attainment age is calculated by dividing the total score by 16. Overall, the test is easy to administer and the instructions are extremely clear.

Technical Data: None.

Limitations: The PAR is a research instrument; consequently it needs further refinement. No normative data are given, not are any reliability or validity data presented. The item order is simply based on a series of small-scale studies. Finally, it would be helpful to have examples of the behaviors in question in addition to the definitions.

Considerations: The Preschool Attainment Record has the advantage over conventional tests such as the

Stanford-Binet that it assesses motor skills, social competencies, and creativity, as well as intellectual abilities. On the whole, the PAR appears to be a valuable instrument for evaluation of handicapped preschoolers. The norming, reliability, and validity data, when added, should increase the value of this test.

RAVEN's PROGRESSIVE MATRICES

Purpose: The Raven's Progressive Matrices is designed to measure the ability to understand nonsense figures, see relationships between the figures, and then imagine what a missing figure would be like. Through the accomplishment of these functions, the test seeks to develop a systematic method of reasoning, Additionally, it meaures and individual's ability to observe as well as think.

Organization: The scale is designed for use with individuals ages 5 years to adult. It consists of 60 problems divided into five sets (A, B, C, D, and E) of 12 items each. In each set, the first problem is (as nearly as possible) self-evident. The problems that follow become progressively more difficult. The five sets provide opportunities for grasping the method of working and five progressive assessments of a person's capacity for intellectual activity. The figures in each problem are boldly presented, accurately drawn, and pleasing to look at. The scale is intended to cover the entire range of intellectual development from the time

a child is first able to grasp the idea of finding a missing piece to the time he is able to complete a problem.

Administration: Every individual, regardless of age, is given an identical series of problems in exactly the same order. The scale is not timed, and the individual works without interruption from the beginning to the end of the scale. The length of the scale is designed so as to assess a person's maximum capacity to form comparisons and to reason by analogy without being exhausting. The scale can be administered as either an individual or group test.

Technical Data: The Raven correlates with the WISC over all age levels from .27 to .91 with a median correlation of .51. The median correlation with the WPPSI is .30.

Limitations: The scale does not differentiate very clearly between young children or between adults of superior intelligence. Additionally, the manual for the Raven Scale is quite inadequate, giving little information on reliability and none on validity. Since this information is so inadequate, the Raven

can only be used as a second test to substantiate another intelligence scale.

Considerations: Due to its nonverbal nature, the test is well-adapted for use with hearing impaired or culturally deprived children. Furthermore, there are two forms, the board form and the book form, that can be used with the orthopedically handicapped since they require only a pointing response.

SANTA CLARA INVENTORY OF DEVELOPMENTAL TASKS

Purpose: The Inventory of Developmental Tasks (IDT) was developed as an E.S.E.A. Title I project of the Santa Clara Unified School District in 1967. The 60 items on the scale were designed to index those milestones in a child's development that are prerequisite to success in more advanced cognitive tasks.

Organization: The items are sequenced by chronological age: preschool, 5 to 5½ years, 6 to 6½ years, and 7 years. The scale assesses development in eight different areas: motor coordination, visual motor performance, visual perception, visual memory, auditory perception, auditory memory, language development, and conceptual development.

Administration: No background in psychological testing is required to administer the scale. The IDT is designed to aid in individualizing instruction for those children with developmental lags in the eight areas. Authors include numerous remediation activities for each item. It is suggested that remediation be begun at the lowest developmental level where the child receives a score of either one or zero.

The child does not receive an overall developmental score. The IDT provides a developmental profile, which is an individual student record on which the performance on each of the 60 items is charted. The tasks become progressively harder and increasingly more abstract as you move from the bottom to the top of the profile chart in each age grouping and within each category as they move from left to right.

Technical Data: This instrument has not been standardized, and there are no norms. Additionally, reliability and validity data have not been provided.

Limitations: Since this scale is frequently administered by individuals with no background in psychoeducational assessment, the obtained results must be interpreted with caution. Also, comprehensive technical information on the validity of the scale, reliability of the items, and standardization procedures is required.

Considerations: The IDT is extremely useful with both young children and older developmentally delayed children. Since there are no norms, it is possible to obtain useful information on children older than the age

range of the scale. It is not necessary to administer all of the items or even administer them in the listed sequence. Therefore, this scale is ideal for "picking and choosing" only those areas where information is wanted.

Since the list is essentially criterion-referenced, it yields information as to whether a child can do a certain task and adapts nicely to use with different low incidence populations. Instruction on mnay of the items are very simple; finger spelling, pantomime, or written instructions could be used to assess a hearing impaired child on all but the two auditory areas. Even on language and conceptual development, a hearing impaired child could communicate his response without being able to speak, via finger spelling or written responses. A visually impaired child could be assessed on the two auditory areas and on motor coordination. The scale could also be used with certain orthopedically handicapped children to assess the degree of their handicap before selecting an intelligence scale.

SEQUENTIAL TESTS OF EDUCATIONAL PROGRESS (STEP)

Purpose: STEP was developed by the Cooperative Test Division of the Educational Testing Service. It represents a major effort to construct objective items measuring understanding rather than mere recall of information. The test focuses on the application of learned skills in the solution of new problems, verbal comprehension, communication skills, and on the outcomes rather than the contents of school learning.

Organization: These test are available in four levels, suitable for grades 4 to 6, 7 to 9, 10 to 12, and 13 to 14. At each level there are six multiple-choice tests in reading, writing, mathematics, science, social studies, listening, and an essay test. Two parallel forms of each objective test and four parallel forms of the essay test are available for each child.

Administration: Although they are timed, the STEP tests are essentially power tests. The essay test requires 35 minutes and the six objective tests require 70 minutes each.

Technical Data: Raw scores on the STEP batteries can be converted into percentile bands that encompass a distance of one standard error of measurement

on either side of the corresponding percentile. Each test is long enough to insure satisfactory reliability. One of the strong points of the STEP tests is the content validity. Committees of educators, representing all grade levels and chosen in consultation with national professional organizations, participated with test construction specialists both in drawing up test specifications and in preparing and reviewing items.

Limitations: The time required to administer the entire battery can be quite lengthy. Consequently, on occasion, steps may need to be taken to compensate.

Considerations: Through a careful and technically complex process, STEP has been adapted for use with blind students. This adaptation appears to be quite sound; consequently, confident use of this test with blind children appears to be justified.

SLOSSON INTELLIGENCE TEST

Purpose: The Slosson Intelligence Test is a short individual intelligence test that is useful as a screening instrument. It has been designed for use by teachers, principals, psychologists, social workers, and other professionals who often need to evaluate an individual's mental ability.

Organization: The majority of items on the Slosson are similar in nature to the Stanford-Binet. Items that are used in the infant and early childhood section were taken directly or adapted from the Gesell Developmental Schedules.

Administration: The test is relatively easy to administer. Testing time is only 15 to 30 minutes, and the obtained raw score can be computed into a mental age and ratio IQ.

Technical Data: High reliability coefficients have been generated for the Slosson. The test correlates highly with the Stanford-Binet, but much lower correlations are obtained when it is compared with the Wechsler full scale.

Limitations: Although the authors did attempt to obtain a representative sample, population norms are not

well described. Furthermore, the Slosson can never be used as a substitute for a comprehensive test of intelligence (i.e., Wechsler Intelligence Scale for Children or Stanford-Binet Intelligent Scale). It is a screening device and nothing more. Finally, the test must be used with caution with children who have delayed languaged development and with those who are not familiar with middle-class language patterns.

Considerations: The Slosson can be adapted for use with handicapped individuals. For example, when testing the blind, the examiner can omit those inappropriate items that require visual abilities. When those items are omitted, the examiner must then use his clinical judgement in scoring that particular level. Additionally, the test can be given to a deaf person if that individual can read. To do this, the examiner must show the individual the questions and cover up the answers.

STANFORD ACHIEVEMENT TEST

Purpose: The Stanford Achievement Test is perhaps the most widely used group-administered achievement test. It was designed to measure academic competence in general subject matter areas at both the elementary and high school grade levels.

Organization: At the elementary school level, the Stanford Achievement Test seeks to measure the student's level of academic functioning in reading, science, social studies, spelling, and language. The high school battery focuses on measuring acquired academic skills in English, spelling, mathematics, numerical competence, reading, science, social studies, and technical comprehension.

Administration: This test is usually administered to groups of students; however, it can also be administered to children individually. The tests are attractive, and administration is facilitated by the color coding of the test booklets and protocols at each level. IBM answer sheets are provided for computer scoring.

Technical Data: The High School Battery was standardized on 22,699 students in 58 schools representing nine geographical regions. The reliability and validity

data are sufficiently adequate for the user to employ this portion of the test with confidence. Standard scores (in terms of grade scores) are given for each test, along with split-half and Kuder-Richardson reliability estimates. Generally, the reliability data seem to be adequate and compare favorably with reliability data for similar batteries.

Limitations: The Elementary School Battery has one major weakness in its norming: there are no local or regional norms. Additionally, low reliability coefficients have been calculated for several of the individual tests (.70 to .60), thus making it inadvisable to employ individual tests without the rest of the battery. Furthermore, the user is not given any direct empirical evidence on the stability of tests within the battery or of the battery as a whole.

Considerations: Forms X and W of the Stanford Achievement Tests have been adapted for use in braille or large print form. A major investment of time and effort has gone into this adaptation and it has been very successful. The only disadvantage is that the time

required to administer the braille edition is considerably greater than the time required to administer the regular print versions with nonvisually handicapped children.

STANFORD-BINET INTELLIGENCE SCALE: 1972 NORMS EDITIONS

Purpose: The Stanford-Binet Intelligence Scale was originally developed by Alfred Binet (who is frequently called the father of intelligence testing) in 1905 to assist the Paris Public School System in their attempt to separate mentally retarded from non-mentally retarded children. The original scale has been revised three times, in 1916, 1937, and 1960, with new norms having been established for the third revision in 1972. Today, the Stanford-Binet, along with the Wechsler Intelligence Scale for Children-Revised, is generally recognized as one of the most frequently used measures of mental ability. Although the instrument has been standardized for use with adults, its primary purpose is to assist in determining the academic potential of children.

Organization: Binet considered judgment, comprehension, and reasoning to be the essential components of intelligence. His basic assumption in designing the scale was that "a person is thought of as normal if he can do the things persons of his age normally do, retarded if his test performance corresponds to the

performance of persons younger than himself, and accelerated if his performance level exceeds that of persons his own age." Therefore, he designed an age scale and employed the concept of Mental Age (MA) as a means of ranking individuals. Terman, who helped revise the scale in 1916, first employed the term IQ by computing the ration of test score to chronological age.

The 1960 scale is a single scale where the best subtests from the two forms of the 1937 scale were used. The subtest were selected so as to make an "arrangement of tests to that the average mental age that the scale gives agrees closely with chronological age." The difficulty and locations of the tests were determined by the percentage passing at each successive MA level, as checked against the 1937 scales. Internal consistency was determined by the correlations of each subtest with the total score. The third revision uses deviation IQs for ages 2 through 18. A deviation IQ as opposed to a ratio IA indicates the same relative ability at different ages.

Technical Data: The 1969 scale was restandardized because it was felt that the original norming sample did not accurately reflect the 1960 population of the United States (the early sample included only white, English-speaking subjects). The new norms used a representative sample of approximately 100 subjects at each age level (approximately 2100 total). The stratifying variable was the test scores of the verbal battery of the Cognitive Abilities Test (which was being standardized at that time), thus resulting in a proportional representation of ability levels. The sample selected included children with Black and Spanish surnames, although the primary language was English.

Biserial correlations were computed for each subtest of the 1937 scale for both forms L and M. Retention of a test for the 1960 scale depended upon its correlation with the total score. The mean correlation for the subtests of the 1960 scale was .66. Comparison with the total score shows that verbal tests were higher (mean .65) than nonverbal test (mean .58). The choice of items according

the MA on the 1937 scale assures that the new scale is measuring the same thing as measured by the original scale. There is a regular increase in MA from one age to the next which corresponds to an increase in percentage passing from one chronological age MA to the next on both forms of the 1937 scale. Many studies of validity have been conducted on the Binet. Overall, it appears to be a good, valid test of intelligence, as defined by its authors.

Limitations: Sattler (1974) lists the following criticisms of the Stanford-Binet: (1) it places too heavy an emphasis on verbal and rote memory tests, (2) there are too few tests of G (general intelligence), (3) there is only one score (IQ) to represent the complex nature of cognitive functions, (4) the scale fails to measure creative abilities, and (5) the scale is unsuitable for testing adults.

Comments: The fact that the Binet is heavily loaded with verbal items makes it useful in testing visually impaired and orthopedically handicapped individuals. When assessing youngsters with these handicapping

conditions, the upper levels of the scale are more adaptable in that they rely less heavily on performance items than the lower levels. In these instances, it is also possible to omit some performance items and substitute the alternate item at a given level.

VINELAND SOCIAL MATURITY SCALE

Purpose: The Vineland Social Maturity Scale is a developmental schedule concerned with an individual's ability to look after his practical needs and to participate in those activities that lead toward independence as adults.

Organization: The items on the scale are arranged in order of increasing difficulty and represent progressive maturation in self-help, self-direction, locomotion, occupation, communication, and social relations. Although it covers the range from birth to over 25 years, the scale is most useful at the younger age levels and with lower functioning individuals.

Administration: The entire scale consists of 117 items grouped into year levels. The information required for each item is obtained through an interview with an informant or with the subject himself. The scale is based on what the subject has actually done in his daily living. The manual provides detailed descriptions and examples of each item to insure correct responses. A social age (SA) and a social quotient (SQ) can be computed from the subject's record on the entire scale.

Technical Data: The scale was tentatively standardized on 620 subjects, including 10 males and 10 females at each year from birth to 30 years. Test validity was determined chiefly on the basis of age differentiation, comparison of normal and subnormal subjects, and correlation of scores with judgments of observers who knew the subject well. High test-retest reliability coefficients are reported. Finally, correlations between the Vineland and Stanford-Binet vary widely, but are low enough to indicate that different facets of behavior are being tapped by the two scales.

Limitations: The norms available for the Vineland must be regarded as tentative since the standardization sample included too few cases at each age level to insure desirable representativeness. Moreover, many items (such as those relating to "going out along," use of "spending money," etc.) would have different significance for children in different socioeconomic levels or minority groups. Furthermore, cultural differences in child-rearing customs, rather than child's ability level, might in some

cases account for deviations from the norms. Similarly, a considerable number of items are unsuitable for institutionalized children.

Considerations: Although its specific limitations are readily apparent, the Vineland can be used effectively with handicapped children. Finally, it is particularly good as an assessment instrument in situations where traditional assessment tools are not functional.

WECHSLER INTELLIGENCE SCALE FOR CHILDREN--REVISED

Purpose: The Wechsler Intelligence Scale for Children--Revised (WISC-R) was designed to measure intellectual ability and academic potential in children ages 6 to 16-11. In designing this instrument, Wechsler sought to operationalize intelligence as "the overall capacity of an individual to understand and to cope with the world around him." Furthermore, he specified that he sees intelligence as manifesting itself in many forms, thus requiring that an intelligence scale include many different subtests if intelligence is to be measured effectively.

Organization: The WISC-R includes 12 different subtests which are divided into two different scales, a verbal and a performance scale. It is only necessary to give any 10 of the subtests to estimate an individual's ability. It is important to note that Wechsler does not consider each of the subtests equally as good as the others in measuring intelligence, but each is considered necessary for a full appraisal of intelligence.

Recognizing that the use of mental age as a basic measure of intelligence is often confusing and misinterpreted, Wechsler does not use the concept of mental age. Instead, he uses the deviation IQ. By using standard deviations that are identical from year to year, and IQ score will not vary unless the actual test performance, as compared with the child's peers, varies.

Wechsler's original 1949 scale was revised in several ways for the 1974 version. First, the age norms were changed so that there would be less overlap with the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) and so that it would be more useful with high school children. Furthermore, the revised scale was normed to include a proportional representation of nonwhites. Finally, those itmes of the 1949 scale which were felt to be obsolete or differentially unfair to particular groups of children were either modified or eliminated.

Administration: Although the WISC-R has been standardized for administration to children as young as 6 years old, it is generally recognized as providing the most

valid estimate of intelligence if administered to children who are at least 8 years old. The WISC-R must be administered to students individually. Also, instructions for administrating and scoring each subtest are very specific and must be followed carefully if obtained scores are to be recognized as valid.

Technical Data: A split-half procedure (odd vs. even items) was used to measure the internal consistency of the WISC-R. The average coefficients across the age levels range from .70 to .86 on the subtests. Reliability coefficients for verbal, performance, and pull scale IQ's were obtained by a formula for computing the reliability of a composite group of tests. The average coefficients are: verbal .94, performance .90, and full scale .96. To test the stability of the WISC-R, 303 children were retested after an interval of about one month. They were divided into three broad groups. For these groups, the average coefficients ranged from .65 to .88.

To determine validity, the WISC-R was compared with three other individually administered scales. The full scale of the WISC-R and the full scale score of the WPPSI correlated at .82; the full scale score of the Wechsler Adult Intelligence Scale (WAIS) at .95. Finally, the correlations with the Stanford Binet, Form LM, range from .71 for verbal to .60 for performance and to .73 for the full scale scores.

A stratified sampling plan was adopted to insure that the normative sample would include representative proportions of various classes, based on the 1970 United States Census. The variables used to make the selections of who would be in the sample include age, sex, race, geographical region, occupation of the head of the household, and urban-rural residence. There were 200 children in each of the 11 age groups from age 6½ to age 16½ years. Therefore, the total sample included 2200 children. There were 100 boys and 100 girls at each level. The United States was divided into four major geographical regions: North East, North Central, South, and West. Nonwhite populations included Blacks,

American Indians, Orientals, Puerto Ricans, and Chicanos. The sample was limited to "normal" children, thereby eliminating the institutionalized mentally retarded and severely disturbed children from the norming populations.

Limitations: Although the WISC-R readily lends itself to modification for use with various low incidence populations, caution is needed when interpreting the results, since it is not known how modifications affect the reliability and validity of the scale. Therefore, the resulting scores should be considered as rought estimates of intelligence.

Considerations: The format of the WISC-R (performance and verbal subtests mixed) makes it enjoyable for most children. The Revised scale is designed to be easier to administer and score than the old scale. Furthermore, the instrument is an excellent predictor of success. Finally, the scale's construction facilitates the modification of certain administrative procedures, thus permitting its use with a variety of handicapped populations. However, the results must be interpreted cautiously. (See Sattler (1974) for selected modification procedures for consideration.)

References

Amos, H. Ontario School Ability Examination. Toronto: The Ryerson Press.

Bayley, N. Bayley Scale of Infant Development. New York: The Psychological Corporation.

Boehm, A. Boehm Test of Basic Concepts. New York: The Psychological Corporation.

Brownley, J. Unpublished research comparing WISC-R with the Leiter International Performance Scale.

Burgemeister, B., Blum, L., & Lorge, I. Columbia Mental Maturity Scale: Guide for Administering the Interpreting. Harcourt, Brace, Javanovich, Incorporated, 1972.

Doll, E. A. Vineland Social Maturity Scale, Circle Pines, Minnesota: American Guidance Service, Inc., 1965.

Doll, E. A. Preschool Attainment Record, Research Edition. American Guidance Service, Inc.

Dunn, L. M., Markwardt, F. C., Jr. Peabody Individual Achievement Test. Circle Pines, Minnesota: American Guidance Service, Inc., 1970.

Dunn, L. M. Peabody Picture Vocabulary Test. Circle Pines, Minnesota: American Guidance Service, Inc., 1965.

Frankenburg, W. & Dodds, J. Denver Developmental Screening Test, Ladoca Project and Publishing Foundation.

French, J. L. Pictorial Test of Intelligence. Boston: Houghton Mifflin, Co., 1964.

Gainer, W., Zweig, R., Dole, P., & Wah, S. Santa Clara Inventory of Developmental Tasks. Santa Clara, California: Richard Zweig Associates, 1974.

Gesell, A. Gesell Developmental Schedules. New York: The Psychological Corporation.

Harris, D. B. Goodenough-Harris Drawing Test Manual. New York: Harcourt Brace, and World, Inc., 1963.

Hiskey, M. Hiskey-Nebraska Test of Learning Aptitude. Union Press.

Kelly, T., Madden, R., & Gardner, E. Stanford Achievement Tests. Harcourt, Brace, Javanovich. Braille and large type editions available from American Printing House for the Blind.

Leiter, R. G. General Instructions for the Leiter International Performance Scale. Chicago: Stoelting Company, 1969.

Nihira, K., Foster, R., & Leland, H. Adaptive Behavior Scales, American Association on Mental Deficiency.

Ravens, J. C. Progressive Matrices. New York: The Psychological Corp.

Saettler, J. Assessment of Children's Intelligence. Philadelphia: W. B. Saunders, 1974.

Shurrager, H. & Shyrrager, P. Haptic Intelligence Scale for Adult Blind. New York: The Psychological Corporation.

Slosson, R. Slosson Intelligence Test for Children and Adults. East Aurora, New York: Slosson Educational Publications, 1975.

Terman, L. M. Stanford-Binet Intelligence Test for Children and Adults. Boston: Houghton Mifflin Company, 1973.

Vernon, M., & Brown, D. W. A guide to psychological tests and testing procedures in the evaluation of deaf and hard-of-hearing disorders. Journal of Speech and Hearing Disorders. Journal of Speech and Hearing Disorders. November, Vol. 29, No. 4, pp. 415-23, 1964.

Wechsler, D. Wechsler Intelligence Scale for Children--Revised. New York: The Psychological Corporation, 1974.

Wieland, M. Psychoeducational Diagnosis of Exceptional Children. Springfield, Illinois: Charles Thomas, 1974.

Yarnall, G. D. Preferred methods of communication by adult deaf-blind persons: A field study of selected individuals. (Accepted for publication by Journal of Rehabilitation for the Deaf. n.d.)